

# Judgement [supplementary chapter]

## Assessing without levels

*This is a personal account, by Mary James, of the reasoning behind a major innovation in the assessment of learning in English schools. The dilemmas in developing methods of assessment which are educationally constructive, by providing feedback and enhancing motivation, are apparent.*

## Introduction

In England, before the introduction of the 2014 National Curriculum, assessment of pupil learning was based on teacher judgement of the 'best fit' between performance and pre-specified 'level descriptors' in each subject and in each year. Teachers generally found this to be workable, but there was also evidence that the system could lead to categorisation of children and distortion of the curriculum.

For the 2014 curriculum, Ministers decided to abandon any requirement for schools to use levels to categorise student attainments. Schools were invited to develop new ways of assessing learning as it takes place. National, statutory assessments were to be retained only at the end of each Key Stage.

As a member of the Expert Panel to the National Curriculum Review in 2011, I was implicated in this decision because we argued that, instead of 'narrowing the gap' between the advantaged and the disadvantaged, which the Government claimed to be concerned about, there was evidence that the existing system exacerbated social differentiation. Moreover:

[The levels system] distorts pupil learning, creating the tragedy, for instance, that some pupils become more concerned for 'what level they are' than for the substance of what they know, can do and understand. (NCR EP Report, para 8.4)

The Expert Panel concluded that:

...all assessment and other processes should bring people back to the content of the curriculum (and the extent to which it has been taught and learned), instead of focusing on abstracted and arbitrary expressions of curriculum such as 'levels'. ... Summary reporting in the form of grades and levels is too general to unlock parental support for learning, for effective targeting of learning support, and for genuine recognition of the strengths and weaknesses of schools' programmes.

(NCR EP Report, para. 8.24, my emphasis)

I stand by these conclusions and believe that the decision to get rid of levels was the right thing to do. But I recognise that this challenges schools to develop alternatives.

The key criterion most often quoted by assessment experts is 'fitness for purpose'. This should also be in the forefront of teachers' and school leaders' minds as they seek to create their own approaches to assessment, recording and reporting. But before getting down to details, it requires them – or should require them – to consider fundamentals and think seriously again about purposes, and how these are grounded in educational values and principles.

In all discussions of data, indicators, targets and impact, it is easy to forget that education is not merely a technical activity for economic ends i.e. to get young people into work and to promote the competitiveness of UK plc. Essentially, education is an ethical activity with moral purpose to enhance human lives in all its aspects – economic, yes, but also spiritual, moral,

cultural, mental and physical. This is what the 2002 Education Act requires. At every level, therefore, deliberation about curriculum and assessment involves judgement, not just measurement.

## Political versus professional imperatives

You won't find much discussion by politicians of the moral purposes of education. Unfortunately, whatever their party, they seem to have become obsessed with accountability and the need for data to allow more discrimination between students (for selection), teachers (for performance monitoring), schools and local authorities (for funding and control decisions).

The driver for this seems to be the belief that higher measured scores on national and international indicator systems will raise the country's economic competitiveness in a global market. Numerical scaled scores in end of Key Stage tests, and finer grading at GCSE, allow more differentiation to identify 'the best' and 'the worst' and convince the public that the consequences are fair because the measures are 'objective'.

In 1988, the Task Group on Assessment and Testing (TGAT) wrote a report which first framed National Curriculum Assessment. This recognised four purposes of assessment and testing – formative, diagnostic, summative and evaluative – but it also argued that the formative purpose should have priority before the age of 16 because of its power to directly support learning itself. Over time however, it is the 'evaluative' purpose that has come to dominate.

Michael Gove was the Secretary of State for Education when the 2014 National Curriculum was being developed. In his statement<sup>2</sup> following his speech to the NCTL conference in June 2013, he said that 'Schools will be able to introduce their own approaches to formative assessment, to support pupil attainment and progression'. But he moved swiftly on to talk about 'pupil tracking data'. In this statement, as in many others that preceded and followed it, I am left with the strong impression that formative assessment is only understood in policy circles (and elsewhere unfortunately) as frequent, mini-summative assessments for recording, reporting and management.

So we need to remind ourselves of the widely-endorsed, research-based definition of Assessment for Learning<sup>3</sup> (or formative assessment) developed by the Assessment Reform Group, of which I was a member, in 2002:

Assessment for Learning is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there. (ARG, 2002)<sup>4</sup>

Underpinning this definition are ten principles:

### *Assessment:*

- *Is part of effective planning*
- *Focuses on how pupils learn*
- *Is central to classroom practice*
- *Is a key professional skill*
- *Is sensitive and constructive*
- *Fosters motivation*
- *Promotes understanding of goals and criteria*
- *Helps learners know how to improve*
- *Develops the capacity for self-assessment*
- *Recognises all educational achievement*

(ARG, 2002)

As teachers and school leaders, our concern should be that our assessment systems, as for all school systems, should first and foremost serve the learning of our students. That is what the business of education is supposed to be about.

Setting schools free to choose how they assess learning as it develops week-by-week over the years, opens up a space for professional innovation in line with these principles. Of course, it also enabled the Coalition Government of the time to claim to give schools more autonomy, and to fulfil its promise to reduce government and save costs. However, 'setting schools free' can only work if overall accountability demands (and their expression in league tables and Ofsted) don't dominate and kill off principled innovation.

Teachers have got to believe, and demonstrate, that doing what is truly educational will genuinely raise standards. Values and practices go together. The research evidence says that principled practice will raise standards – but, through the pressure of external constraints, some teachers have lost confidence in the power of their expertise. This needs to change.

## Implications for assessing without levels

So what has all this got to do with assessing without levels?

Well, it means that formal assessments and tests, and their results, need to be put firmly in their place. Curriculum, teaching and learning (pedagogy) must take centre stage. 'Teaching to the test' is not necessarily wrong, if the test is truly reliable, valid and useful for diagnosing problems in learning. Tim Oates makes an argument that we could benefit if teachers had more access to well-developed banks of assessment items and activities for teachers to use at appropriate points in their teaching of specific and tricky concepts.<sup>5</sup> But, on the whole, formal tests and examinations only provide a 'dipstick' of performance - on limited occasions, on selections from the subject domain. For example, grammar, spelling and punctuation, and reading tests do not capture all, or even very much, of what it is to learn English. Worst of all is when 'teaching the tests' becomes the de facto curriculum and too much time is spent practising for them.

Robert Coe, in his inaugural professorial lecture at Durham University in 2013, claimed that through the assiduous efforts of teachers to prepare students for public examinations, performance scores have risen but real standards haven't. In other words, he believes that grade inflation is a reality. He compared the exponential growth in GCSE scores (a hockey stick graph) with our results in international assessments (where we have flat-lined) and concluded that:

the two sets of data tell stories that are not remotely compatible. Even half the improvement that is entailed in the rise in GCSE performance would have lifted England from being an average performing OECD country to being comfortably the best in the world. To have doubled that rise in 16 years is just not believable. (Coe, 2013, p. v)<sup>6</sup>

In other words, just because students have attained a test result, does not always guarantee that they have actually learned anything. This was nicely illustrated when George Osborne, then Chancellor of the Exchequer, was asked by a seven-year-old boy what seven times eight equals.<sup>7</sup> He refused to answer saying, 'I've made it a rule in life not to answer a load of maths questions'. Tellingly, in response to Sam's previous question, asking whether he was good at maths, he said, 'Well, I did maths A-Level so I have been tested at school'. So he passed the A-Level exam but hadn't learned his times tables – or so it seems!

David Hogan (2014), former Principal Research Scientist at the National Institute in Education in Singapore, analyses the success of Singapore schools in the OECD tests, but also argues that Singapore cannot rest on its laurels.<sup>8</sup> He has advice for us too:

Singapore's experience and its current efforts to improve the quality of teaching and learning do have important, if ironic, implications for systems that hope to emulate its success.

This is especially true of those jurisdictions – I have in mind England and Australia especially – where conservative governments have embarked on ideologically driven crusades to demand more direct instruction of

(Western) canonical knowledge, demanding more testing and high stakes assessments of students, and imposing more intensive top-down performance regimes on teachers.

In my view, this is profoundly and deeply mistaken. It is also more than a little ironic given the reform direction Singapore has mapped out for itself over the past decade. The essential challenge facing Western jurisdictions is not so much to mimic East Asian instructional regimes, but to develop a more balanced pedagogy that focuses not just on knowledge transmission and exam performance, but on teaching that requires students to engage in subject-specific knowledge building.

Knowledge building pedagogies recognise the value of established knowledge, but also insist that students need to be able to do knowledge work as well as learning about established knowledge. Above all, this means students should acquire the ability to recognise, generate, represent, communicate, deliberate, interrogate, validate and apply knowledge claims in light of established norms in key subject domains.

This underlines the importance, in the 21st century, of considering carefully what and how to teach and harmonising this with what and how to assess. Alignment between the two is crucial. Without alignment the latter (assessment) subverts the former (curriculum and pedagogy).

Routine school-based assessments should be referenced to the curriculum - the school curriculum as well as the national curriculum - but not directly to the national tests. If students have truly learned what is in the curriculum, they should do well in the tests and examinations. It should not be necessary to drill them - sacrificing learning to just 'getting a grade'. We all know that grades are important for progression into higher and further education and employment but if current practices were really effective we would not have so many complaints from universities and employers about the quality of recruits and the need for 'remedial' action.

## Aligning assessment to the curriculum

Herein lies the case for putting aside numbers (levels, grades and scores) and capturing teachers' judgements in school-based assessment systems – and developing systems for enhancing and assuring the quality of these judgements. This might be done, for instance, through collaborative moderation processes based on scrutiny of examples of students' work – their actual learning outcomes.

In the early years of National Curriculum Assessment I was involved in a project across the six counties in East Anglia to develop such group moderation systems for KS1 teacher assessment.<sup>9</sup> Within a short space of time, teachers developed procedures and ways of making valid and reliable judgements. But most importantly this innovation developed their understanding of what counts as quality learning and performance. Further, it enhanced their expectations of students. Unfortunately, union action on workload encouraged the then Government to introduce tests at KS1, which effectively killed off this development.

Judgements are best communicated qualitatively, in words, describing achievements and providing advice on improvement, with close reference to the content of the curriculum.

Qualitative judgements can engage students, parents and other teachers in dialogue about the substance of what has been learned, what needs to be learned, and how students can be helped. Numbers do not possess this power, nor do generalised scales disguised as 'performance descriptors'.

Of course, parents are anxious to know whether their child is 'on track', needs help, or exceeds expectations but this comparative information can be communicated by, at most, a three-category system. For example, the Early Years Foundation Stage profile has used 'emerging', 'expected' or 'exceeding'. In truth, this is as fine a comparative judgement that can reliably be claimed for any individual student at any given time. Significantly a number of the packages exemplifying assessment without levels, supported by the DfE Innovation Fund in 2014<sup>10</sup>, adopted a three- or four-category system.

It may be appropriate to create descriptions of what constitutes 'expected' achievement (or what is considered the 'national' standard). But if a student has not achieved that standard and is not ready to progress to the next topic/theme/skill/concept, they, their teachers and their parents need to know in what particular aspects they are struggling and how they can be helped to improve. Likewise, if they have exceeded expectations, they need to know in what particular respects and how they can enrich and widen their achievements. Such extra opportunities should focus on securing and enhancing breadth and depth e.g. through applications of knowledge, rather than racing on to the next curriculum topic.

A three-fold categorisation enables teachers to record and report achievement in any given element of the curriculum in synoptic ways, but it also encourages them to expand, in words, on the help students need if they are struggling, or the extra opportunities required by those who exceed expectations. To create further categories - e.g. 'mastery', 'above national standard', 'national standard', 'working towards national standard', 'below national standard' (as suggested for KS2 Writing) - just creates a grading system by any other name, and is likely to encourage labelling and its perverse consequences.

Of course, progression is as important as absolute attainment, but judgements about progression do not require a generalised ladder, as in the levels system. Progression should be embodied in the curriculum i.e. how content is selected and ordered to reflect not just the logic of the subject but what we know about how the subject matter is learned (which may or may not follow the logic of the subject). Some subject areas have a considerable body of research to help us in this area, particularly in language, physical development, mathematics and science. In other subjects there is still much to be done and teachers will need to make judgements based on their shared experience.

Understandably, school leaders have been concerned about how Ofsted will inspect schools after the removal of levels. In his letter to head teachers in July 2014, the Chief Inspector wrote<sup>11</sup>:

*As happens now, inspectors will use a range of evidence to judge learning and progress. In particular, they will take account of test/examination results, other assessment information and the standard of pupils' work.*

*However, inspectors will:*

- *spend more time looking at the range of pupils' work to consider what progress they are making in different areas of the curriculum*
- *talk to leaders about schools' use of formative and summative assessment and how this improves teaching and raises achievement*
- *evaluate how well pupils are doing against relevant age-related expectations as set out by the school and the national curriculum (where this applies)*
- *consider how schools use assessment information to identify pupils who are falling behind in their learning or who need additional support to reach their full potential, including the most able*
- *evaluate the way schools report to parents and carers on pupils' progress and attainment and assess whether reports help parents to understand how their children are doing in relation to the standards expected.*

Nothing is said here about numbers or scores; qualitative forms of assessment, recording and reporting, can meet all these points. I would argue that they would also be more transparent.

## What assessment without levels might look like

This argument is all very well, I hear you say, but what might it look like in practice?

What follows is one example. It is not deliberately sought out as 'best practice'. It just happens to be something that was introduced into my local primary school in 2014. Teachers and parents wanted to develop performance of poetry as part of their 'school curriculum'. Poetry per se does not appear as a key area of English in the New National Curriculum at KS1 or KS2, although as an ex-English teacher I always considered students' reading, writing, performing, understanding and

enjoying poetry to be a key component of English teaching. I think Michael Rosen, among others, would agree. There is reference to 'learning to appreciate rhymes and poems, and to recite some by heart' etc. but that hardly satisfies.

The teachers and parents at Fen Drayton Primary School seemed to feel the same so they created a poetry speaking competition as a whole school activity in which all classes, and parents, could be involved.<sup>12</sup> The focus was on discussing, choosing, memorising, dramatizing and performing poems to an audience of children, teachers, parents and judges (note the word 'judge'). A list of age-appropriate poems was created for participants to select from (see Figure 1).

A rubric or set of criteria for judgement was developed and tested (see Figure 2). Guidance ('tips for a great performance') based on these criteria was developed and sent to parents so that they could support teachers in coaching the children. Heats were held; feedback given; opportunities for practice and improvement provided before the public performances, which were then judged.

**Figure 1: Poetry Aloud Competition: Poetry Choices – Theme 'Nature'**

Reception

Poem	Poet
Mud	Shirley Hughes
Wind	Shirley Hughes
Little Bird	Michael Foreman's Mother Goose
Today I saw a little worm	Spike Milligan
The Lion	Jack Prelutsky

Years 1 & 2

Poem	Poet
The Hippopotamus	Ogden Nash
The Caterpillar	Christina Rossetti
The Sandpiper	Robert Frost
The Eagle	Tennyson
A Catch	Mary Ann Hoberman

Years 3 & 4

Poem	Poet
The Muddy Puddle	Dennis Lee

Penguins	Leonard Clark
The Lion	Roald Dahl
The Shark	Lord Alfred Douglas
Anthropods	Mary Ann Hoberman

Years 5 & 6

Poem	Poet
The Crocodile	Roald Dahl
The Shell	James Stephens
The Lobster Quadrille	Lewis Carroll
The way through the woods	Rudyard Kipling
The Tyger	William Blake

Adults

Poem	Poet
Macavity: the mystery cat	T.S. Eliot
April Rise	Laurie Lee
Blackberry Picking	Seamus Heaney
The Thought-fox	Ted Hughes
The Wind King	Tony Layton

**Figure 2: Performing poetry rubric**

	Weak	Competent	Good	Excellent
<b>Physical presence</b>	Timid; unsure; eye contact and body language reflects nervousness	Body language and eye contact are at times unsure, at times confident	Comfortable; steady eye contact and confident body language	Poised; body language and eye contact reveal strong stage presence

<b>Voice and articulation</b>	Inaudible; too loud; monotone; paced unevenly; singsong; hurried; mispronunciations	Clear, adequate intonation, even pacing	Clear, appropriate intonation and pacing	Very clear, crisp, effective use of volume, intonation, rhythm, and pacing
<b>Dramatic appropriateness</b>	Poem is secondary to style of delivery; includes instances of distracting gestures, facial expressions, and vocal inflections; inappropriate tone	Poem is neither overwhelmed nor enhanced by style of delivery	Poem is enhanced by style of delivery; any gestures, facial expressions, and movement are appropriate to poem	Style of delivery reflects precedence of poem; poem's voice is well conveyed
<b>Evidence of understanding</b>	Doesn't sufficiently communicate meaning of poem	Satisfactorily communicates meaning of poem	Conveys meaning of poem well	Interprets poem very well for audience; nuanced
<b>Overall performance</b>	Inadequate recitation; lacklustre; does disservice to poem	Sufficient recitation; lacks meaningful impact on audience	Enjoyable recitation; successfully delivers poem	Inspired performance shows grasp of recitation skills and enhances audience's experience of the poem

This assessment activity had certain notable features:

- The criteria were developed very specifically in relation to this content domain; they were not overly abstract or generalised.
- The same rubric was used for all ages of children (and the adults).
- Progression was embodied in the poems, which became more difficult for the older classes. The criteria had to be interpreted in relation to the specific poems chosen, and performance judged accordingly.
- Four categories were used to discriminate different standards of performance on the criteria. Initially it was thought necessary to attach a numeric score to each of these in order to aggregate into an overall score. But, in practice it was unnecessary and cumbersome to use numbers. The judges (I was one) found it quite possible to come to judgements – and a rank order for awarding prizes - based on the descriptions alone. Three categories would probably have sufficed.
- The response to this activity was very positive; in the end of year reports to parents many children mentioned the poetry competition as the highlight of their year.
- At least one small boy gave the performance of his life (so far) and changed his teachers' assumptions about his ability.

Perhaps the most important point to note is the need to develop criteria for assessment that relate directly to the content area being taught and learned at a given time. If summarising labels are used, the criteria that underpin them must be clear so that what counts as, say, 'weak', 'competent', 'good', 'excellent' – in this specific context - is explicit.

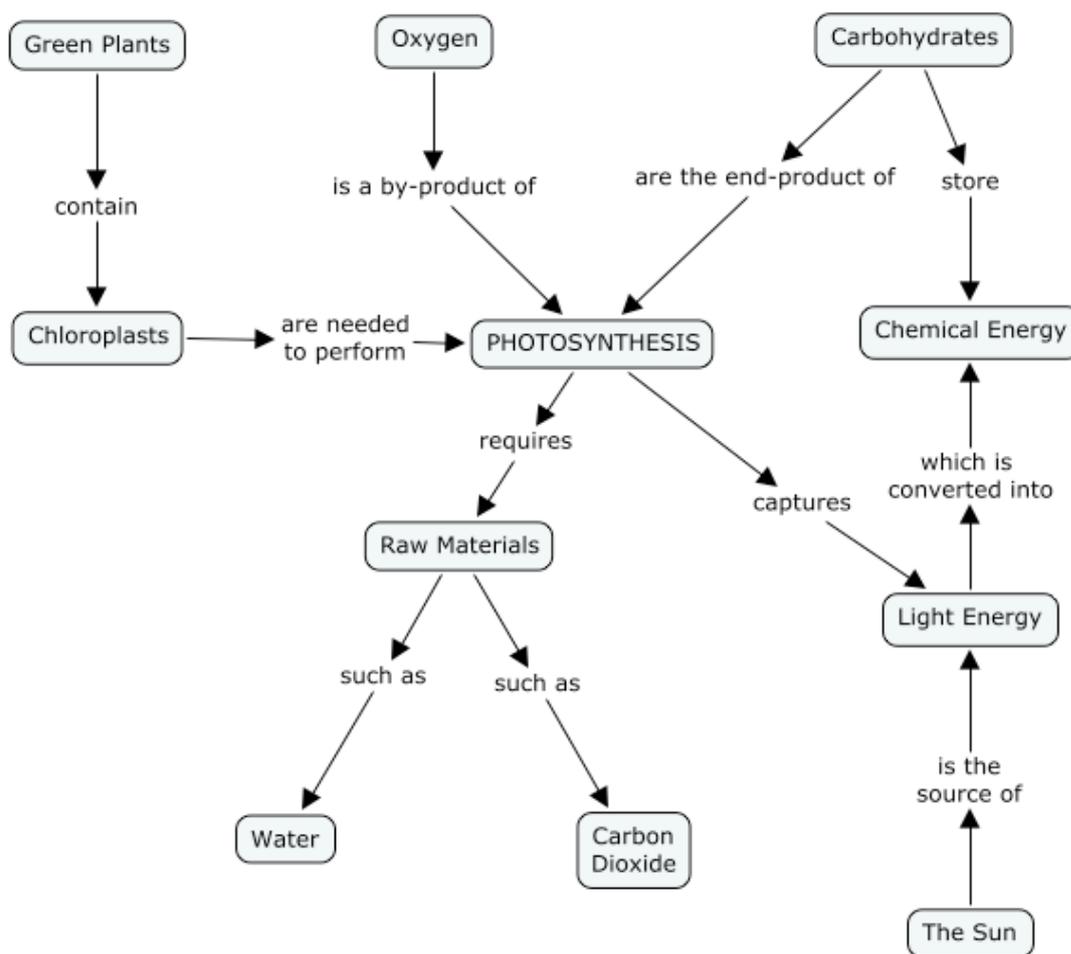
Another example, this time from science, may reinforce this point.

Concept maps can be an important tool for assessing students' knowledge and understanding of big ideas in science. Concept maps consist of nodes (concepts – the nouns) and lines (symbolising relationships – the verbs). Students can be asked to 'construct a map' or 'fill in a map' depending on the teachers' judgement of the degree of challenge that what would be appropriate at the time. The students' maps can then be compared with an 'expert' map and judged diagnostically, formatively and summatively, according to purpose.

Photosynthesis can be regarded as a big idea in science; an expert concept map is found in Figure 3. Properly interrogated, to assure a student's understanding, this can demonstrate end-point achievement. But it is usually important to know the steps that the student has made towards this understanding – their progression.

Other work in science education, particularly by Paul Black and Rick Shavelson and Wynne Harlen, has been focusing on learning progressions in science. In 2006, Harlen formulated some indicators of progression from 'small' to 'big' ideas (Figure 4).<sup>13</sup> However she acknowledged that these are 'generic' statements that need to be 'translated' in the context of learning particular subject content in particular activities. Thus, as in the 'Poetry Aloud' example where the criteria need to be considered alongside the difficulty of the poems, so, in learning photosynthesis in science, the generic criteria need to be put alongside the concept map. The generic can only be properly understood in relation to the specific. The question needs to be asked: 'What would progression, in learning this particular big idea in science, look like?'

**Figure 3: Photosynthesis concept map<sup>14</sup>**



**Figure 4: Indicators of progression in ideas in science**

<p>Students:</p> <p>Offer a description only with no attempt to explain a situation.</p> <p>Use a pre-conceived non-scientific idea to explain a situation or make a prediction.</p>
--

Name a relevant idea (for example, 'friction', 'gravity') without explaining its relationship to the situation.  
Use a relevant idea to explain a specific situation but not other situations where it also applies.  
Use a relevant idea to explain several related experiences or situations.  
Refer to a 'bigger' idea that explains a number of linked phenomena.  
Use a bigger idea to predict events not yet encountered.

(Based on Harlen 2006, p.148)

## Conclusion

It is not reasonable to expect all schools and teachers to develop a full range of unique assessment activities, such as those given above, to cover all the key content areas in the school and national curriculum. But they can build them incrementally – possibly with the help of their communities and publishers - as part of development planning.

To have the incentive to do this, teachers need to believe that Government is genuine in its stated commitment to giving them freedom to innovate and use their professional judgement. If trust in teachers is seen to be half-hearted then 'teaching the test' or 'teaching the performance descriptor' will continue to wash back from end-of-key-stage testing to all intervening years. The 'indicator' of achievement will become the 'target' and have little value as a means of judging what students really know.

In order to avert this kind of thing, we really do need to get rid of the excessively high stakes attached to test results. The drive to collect data for system management is undermining the use of assessment information to improve students' learning. Getting rid of league tables would help, as would letting a re-professionalised Inspectorate (national and local) do its job in assessing how well schools 'understand each student's progress and needs and how clearly they communicate this to students parents and Governors'.

Of course, tests and exams at the end of compulsory schooling will be high stakes for students, but, if students have been learning well in the preceding years, then they should achieve their goals. There is no need for schools to make their own school-based assessments so high-stakes. They can be high quality without this.

What is most important is to get the underpinning values, purposes, principles and procedures in place – and then build from there. This will also provide a reasonable first response to questions from parents and inspectors about what you are doing now that levels have been removed. It will provide a basis for quality assurance.

I am aware how difficult it is to convince some colleagues, some parents and some students about any move away from levels and grades, especially in a world where increasing amounts of numerical data are used on a daily basis. It will be a considerable task to convince some people that familiar ways of grading students' regular schoolwork are not fit for purpose and need to be changed. We will need to accept responsibility and take active steps to educate, not only our students, but also their parents.

We are asking for nothing less than a change in mind-set.

## Notes

[1] [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/175439/NCR-Expert\\_Panel\\_Report.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/175439/NCR-Expert_Panel_Report.pdf)

[2] <http://webarchive.nationalarchives.gov.uk/20130904084116/https://www.education.gov.uk/schools/teachingandlearning/curriculum/nationalcurriculum2014/a00225864/assessing-without-levels>

[3] The term Assessment for Learning has been dropped from the lexicon of the DfE because the New Labour Government took it up in its national strategies, which were promptly abandoned after the 2010 election.

[4] <http://webarchive.nationalarchives.gov.uk/20110809101133/http://assessment-reform-group.org/CIE3.pdf>

[5] [https://www.youtube.com/watch?v=yDYjF\\_bQy4Q](https://www.youtube.com/watch?v=yDYjF_bQy4Q)

[6] <http://www.cem.org/attachments/publications/ImprovingEducation2013.pdf>

[7] <http://www.theguardian.com/politics/2014/jul/03/george-osborne-refuses-boys-maths-test>

[8] <https://theconversation.com/why-is-singapores-school-system-so-successful-and-is-it-a-model-for-the-west-2291>

[9] James, M. (1994) Experience of Quality Assurance at Key Stage 1. In W. Harlen (ed) *Enhancing Quality in Assessment*, (London: Paul Chapman Publishing) 116 – 138.

[10] See [http://community.tes.co.uk/national\\_curriculum\\_2014/b/assessment\\_without\\_levels/default.aspx](http://community.tes.co.uk/national_curriculum_2014/b/assessment_without_levels/default.aspx)

[11] <http://www.ofsted.gov.uk/resources/note-for-inspectors-use-of-assessment-information-during-inspections-201415>

[12] For documentation

see: [http://www.fendraytonprimary.co.uk/index.php?option=com\\_docman&task=cat\\_view&gid=55&Itemid=59](http://www.fendraytonprimary.co.uk/index.php?option=com_docman&task=cat_view&gid=55&Itemid=59)

[13] Harlen, W. (2006) *Teaching, learning and assessing science 5-12* 4th ed. London: Sage

[14] Source: Shemwell, J., Fu, A., Figueroa, M., Davis, R., and Shavelson, R. (2010) *Assessment in Schools – Secondary Science*. In P. Peterson, E. Baker and B. McGaw, *International Encyclopedia of Education*, 3rd Edition. Oxford: Elsevier). Volume 3:308